Exact Solution of a Model DNA-Inversion Genetic Switch with Orientational Control

Paolo Visco, Rosalind J. Allen, and Martin R. Evans

SUPA, School of Physics, The University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom

(Received 4 April 2008; published 11 September 2008)

DNA inversion is an important mechanism by which bacteria and bacteriophage switch reversibly between phenotypic states. In such switches, the orientation of a short DNA element is flipped by a site-specific recombinase enzyme. We propose a simple model for a DNA-inversion switch in which recombinase production is dependent on the switch state (orientational control). Our model is inspired by the *fim* switch in *E. coli*. We present an exact analytical solution of the chemical master equation for the model switch, as well as stochastic simulations. Orientational control causes the switch to deviate from Poissonian behavior: the distribution of times in the on state shows a peak and successive flip times are correlated.

DOI: 10.1103/PhysRevLett.101.118104 PACS numbers: 87.18.Cf, 82.39.−k, 87.16.Yc

Reversible and heritable stochastic switching between two different states of gene expression is a common phenomenon among bacteria and bacteriophage, known as phase variation. Phase variation is often linked to pathogenesis, and may help bacteria survive fluctuating environmental conditions (e.g., a host immune system) [1]. An important molecular mechanism for phase variation is site-specific DNA inversion [2]. Here, a short piece of DNA (the "invertible element") is excised from the genome and reinserted (strand-by-strand) in the opposite orientation by a site-specific recombinase enzyme binding to sequences at the ends of the invertible element. Different states of gene expression correspond to the two orientations of the invertible element ("switch states"). A well-known example is the *fim* genetic regulatory system, which controls the production of type 1 fimbriae in *E. coli*. These fimbriae are important in uropathogenesis [3]. In the *fim* system, the FimE recombinase is produced more strongly when the switch is in the on state than in the off state [4,5]. This phenomenon is known as orientational control.

In this Letter, we present a simple and general stochastic model for a DNA inversion switch with orientational control. We solve this model analytically, allowing us to determine the range of stochastic switching behavior possible for this type of switch. We find that non-Poissonian behavior occurs, resulting in a peak in the probability distribution of time spent in the on state and correlations between successive flips. Such non-Poissonian behavior could have important effects on the population dynamics of switching microbes in changing environments. One key parameter (the concentration of the recombinase not under orientational control) controls the degree to which our model is non-Poissonian; this parameter corresponds to the main point of environmental regulation for the *fim* switch. In contrast, bistable genetic switches such as positive feedback loops [6] or mutually repressing genes [7,8] in general show only Poissonian behavior (exponential waiting time distributions and uncorrelated flips).

The model.—Our model DNA-inversion switch, illustrated in Fig. 1, contains three elements: the invertible DNA element and two types of recombinase enzyme (R_1 and R_2). The invertible element has two possible orientations (the "on" and "off" states). These correspond to alternative patterns of gene expression, leading to different phenotypic states; however, we model here only the core of the switch and not its downstream effects. The switch can be flipped between its two orientations by either of the recombinases. The concentration of recombinase R_2 is assumed to be fixed, while the production of R_1 depends on the switch state: R_1 is produced only in the on state. This feature of the model constitutes its orientational control and leads to its non-Poissonian behavior. The model is represented by the following reaction scheme:

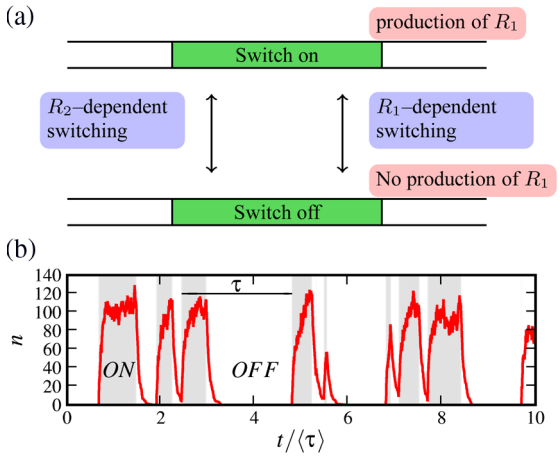

FIG. 1 (color online). (a) Schematic illustration of the model. (b) A typical simulation trajectory. The solid line represents the time evolution of the number n of R_1-molecules, the shading denotes the switch position and τ indicates the duration of a period of the switch. Parameter values are $k_1 = 1$, $k_2 = 100$, $k_3^{\text{on}} = 0.001$, $k_3^{\text{off}} = 0$ $k_4^{\text{on}} = 0.1$, and $k_4^{\text{off}} = 0.1$.

$$R_1 \xrightarrow{k_1} \varnothing \qquad S_{\rm on} \xrightarrow{k_2} S_{\rm on} + R_1 \tag{1a}$$

$$S_{\rm on} + R_1 \underset{k_3^{\rm off}}{\overset{k_3^{\rm on}}{\rightleftharpoons}} S_{\rm off} + R_1 \qquad S_{\rm on} \underset{k_4^{\rm off}}{\overset{k_4^{\rm on}}{\rightleftharpoons}} S_{\rm off}. \tag{1b}$$

Here, $S_{\rm on}$ and $S_{\rm off}$ denote the on and off states of the switch. Reactions (1a) describe the production and decay of recombinase R_1: R_1 is removed from the system with rate constant k_1, and is produced at rate k_2 only when the switch is in the on state. Reactions (1b) describe switch flipping. This may be catalyzed by R_1 with rate constants $k_3^{\rm on}$ (on to off) and $k_3^{\rm off}$ (off to on). We shall mainly consider here the case $k_3^{\rm off} = 0$. Recombinase R_2 can also catalyse switch flipping. The concentration of R_2 (fixed in our model) is implicit through a dependence of $k_4^{\rm on/off}$ on the R_2 concentration.

We note that mean-field, macroscopic rate equations corresponding to the above reaction scheme yield only one steady-state solution corresponding to the average switch state and average concentration of R_1. The underlying deterministic structure of the model is thus not bistable. In this sense, our model is fundamentally different from the bistable reaction networks presented in [6–8].

Our model is inspired by the *fim* genetic regulatory system [3]. In analogy with *fim*, R_1 in our model represents FimE while R_2 represents FimB. Environmental stimuli such as nutrient conditions and temperature act on the *fim* switch largely through changes in the level of FimB [9,10]; this would correspond to variation of our key parameters $k_4^{\rm on/off}$. However, our model is highly simplified in comparison to *fim*, in that it neglects cooperative and competitive recombinase binding and the effects of other DNA binding proteins. Our objective in this work is not to model the details of the *fim* system, as other authors have done [9,10], but rather to address general questions about the behavior of this type of switch. Our model is designed to be as simple as possible while retaining the key features of DNA inversion and orientational control; its simplicity allows us to obtain analytical results and to explore a wide range of parameter space.

We simulated the reaction scheme (1) using a continuous time Monte Carlo scheme [11]. A typical trajectory is shown in Fig. 1, where we plot the number n of R_1 molecules and the switch state as functions of time. When the switch is in the on state, n increases (on average) towards a plateau value of k_2/k_1, while in the off state n decays towards zero. We now obtain an exact analytical solution for the case where $k_3^{\rm off} = 0$. This case is relevant to the *fim* switch, where FimE catalyses almost exclusively on to off switching. In the following, all our analytical results will correspond to $k_3^{\rm off} = 0$, while simulation results will be presented also for $k_3^{\rm off} > 0$. Analytical results for $k_3^{\rm off} > 0$ will be published elsewhere.

Steady state.—We first consider the statistics of n in the steady state and compute the long time, joint probability $p_s(n)$ that the switch is in state s and there are n molecules of R_1. The system of birth-death equations for $p_{\rm on}(n)$ and $p_{\rm off}(n)$ becomes in the steady state

$$(n+1)k_1 p_{\rm on}(n+1) + k_2 p_{\rm on}(n-1) + k_4^{\rm off} p_{\rm off}(n) = (nk_1 + k_2 + nk_3^{\rm on} + k_4^{\rm on}) p_{\rm on}(n), \tag{2a}$$

$$(n+1)k_1 p_{\rm off}(n+1) + nk_3^{\rm on} p_{\rm on}(n) + k_4^{\rm on} p_{\rm on}(n) = (nk_1 + k_4^{\rm off}) p_{\rm off}(n). \tag{2b}$$

In order to decouple the above set of equations, we solve (2a) for $p_{\rm off}$, then insert the result into (2b) to give a decoupled equation for $p_{\rm on}$. Introducing the generating function $G_s(z) = \sum_n z^n p_s(n)$ (where $s = \{\rm on, off\}$), the decoupled equation for $p_{\rm on}$ reduces to a second order differential equation for $G_{\rm on}$. Defining then a new variable $u \equiv u_z = k_2 z/(k_1 + k_3^{\rm on}) - k_1 k_2/(k_1 + k_3^{\rm on})^2$, the latter equation reads

$$uG''_{\rm on}(u) + (a-u)G'_{\rm on}(u) - bG_{\rm on}(u) = 0, \tag{3}$$

where $a = 1 + u_1 + (k_4^{\rm on} + k_4^{\rm off})/(k_1 + k_3^{\rm on})$ and $b = 1 + k_4^{\rm off}/k_1$. Expanding the solution as a regular power series (i.e., $G_{\rm on}(u) = \sum_m a_m u^m$), one finds that $G_{\rm on}(z) = a_0 {}_1F_1(b, a, u)$, where ${}_1F_1$ is a confluent hypergeometric function and a_0 is an integration constant which can be determined with the normalization condition $\sum_n p_{\rm on}(n) + p_{\rm off}(n) = 1$ [or equivalently $G_{\rm on}(1) + G_{\rm off}(1) = 1$]. This result can be rewritten in terms of the original variable n as

$$p_{\rm on}(n) = a_0 \frac{(u_1 - u_0)^n}{n!} \frac{(b)_n}{(a)_n} {}_1F_1(b+n, a+n, u_0), \tag{4}$$

where $(\alpha)_n = \alpha(\alpha+1)\ldots(\alpha+n-1)$. An expression for $p_{\rm off}$ can be derived by inserting Eq. (4) into Eq. (2a). In Fig. 2 we compare the result for $p(n) = p_{\rm on} + p_{\rm off}$ to simulations, obtaining perfect agreement. Figure 2 also illustrates the effects of the different time scales for switch

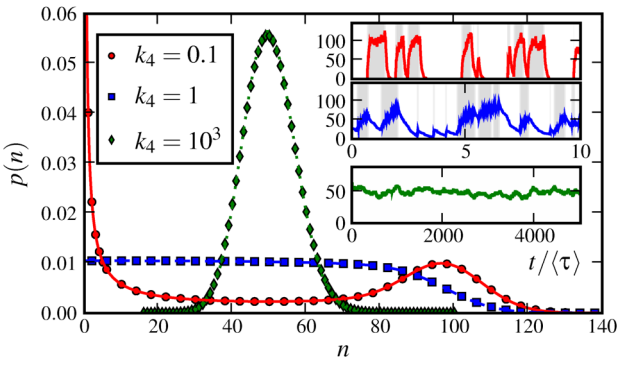

FIG. 2 (color online). Probability distribution for the number n of R_1 molecules, with $k_4^{\rm on} = k_4^{\rm off} = k_4$, for different values of k_4. The symbols show simulation results and the solid lines are the theoretical predictions. For the case where $k_4 = 10^3$, the (dotted) line is a Poisson distribution with parameter $(1 + k_4^{\rm on}/k_4^{\rm off})^{-1} k_2^{\rm on}/k_1$. The other parameters are: $k_1 = 1$, $k_2^{\rm on} = 100$, $k_3^{\rm on} = 0.001$, and $k_3^{\rm off} = 0$. The insets show the typical simulation trajectories for each distribution, where the time is expressed in units of the mean period $\langle\tau\rangle$.

flipping and production/decay of R_1. For small k_4, switch flipping is slow compared to the rate of change of n. In this case, when the switch is in the on state, the number of recombinase has time to reach a plateau before the switch flips off. The on and off switch states are then each associated with a different value of n and the distribution $p(n)$ is bimodal. In contrast, when k_4 is large, the switch flips back and forth much more rapidly than recombinase production or removal. Then, the fraction of time spent in the on state is $k_4^{\text{off}}/(k_4^{\text{off}} + k_4^{\text{on}})$, and $p(n)$ tends to the Poisson distribution expected for a birth-death process with birth rate $k_2^{\text{on}} k_4^{\text{off}}/(k_4^{\text{off}} + k_4^{\text{on}})$ and death rate k_1.

Flipping time distributions.—To determine how orientational control affects switch function, we compute flipping time distributions. The flipping time T can be defined in two different ways. In the first scenario, which we call the switch change ensemble (SCE), we define T as the time spent in a particular switch state—for example, $F_{\text{on}}^{\text{SCE}}(T)$ is the probability distribution for the time between the moment the switch enters the on state and the moment it flips from the on to the off state. In the second scenario, which we call the steady state ensemble (SSE), we start observing the cell at a random moment and measure the time interval between this moment and its next flip into the other state. $F_{\text{on}}^{\text{SSE}}(T)$ and $F_{\text{off}}^{\text{SSE}}(T)$ may be relevant to the response of a population of switching cells to a sudden environmental change. They also correspond to an experiment where one measures the time until the next flip, for cells sampled in the steady state [12]. To compute these distributions, we define $F_s(T|n_0)$ as the probability that the system begins at $t = 0$ in the s state with n_0 recombinase and flips for the first time at T. Note that for $k_3^{\text{off}} = 0$, the off to on flipping process does not depend on R_1 and is governed by k_4^{off}; thus $F_{\text{off}}(T) = k_4^{\text{off}} \exp(-k_4^{\text{off}} T)$ is independent of n_0 (for both the SCE and the SSE). However, the on to off flipping rate is n_0 dependent, so that we average over the ensemble of initial states (characterized by the probability $W_{\text{on}}(n)$ of having n recombinase at the start of our measurement) to obtain the flip time distribution $F_{\text{on}}(T) = \sum_{n_0} W_{\text{on}}(n_0) F_{\text{on}}(T|n_0)$. For the SCE, the initial condition is taken just after a flip, which implies that $W_{\text{on}}^{\text{SCE}}(n_0) = p_{\text{off}}(n_0)/G_{\text{off}}(1)$. For the SSE, the initial condition is sampled in the steady state, yielding $W_{\text{on}}^{\text{SSE}}(n_0) = p_{\text{on}}(n_0)/G_{\text{on}}(1)$. To compute $F_{\text{on}}(T)$, we first define the survival probability $h_{\text{on}}(n, t)$ that, at time t, the switch is in the on state with n R_1 molecules, without having flipped, given the initial condition $h_{\text{on}}(n_0, 0) = W_{\text{on}}(n_0)$. The evolution equation for h_{on} is:

$$\partial_t h_{\text{on}}(n, t) = (n + 1) k_1 h_{\text{on}}(n + 1, t) + k_2 h_{\text{on}}(n - 1, t) - (n k_1 + k_2 + n k_3^{\text{on}} + k_4^{\text{on}}) h_{\text{on}}(n, t). \quad (5)$$

Defining a generating function $\tilde{h}_{\text{on}}(z, t) = \sum_n z^n h_{\text{on}}(n, t)$, it follows that $F_{\text{on}}(t) = -\partial_t \tilde{h}_{\text{on}}(1, t)$. Equation (5) reduces to a partial differential equation for $\tilde{h}_{\text{on}}(z, t)$ which has the initial condition $\tilde{h}_{\text{on}}(z, 0) = \sum_{n_0} W_{\text{on}}(n_0) z^{n_0}$. Its solution is

$$\tilde{h}_{\text{on}}(z, t) = e^{-t(k_4^{\text{on}} + k_2(1 - k_1 \tau_{\text{on}}))} e^{k_2 \tau_{\text{on}}(z - k_1 \tau_{\text{on}})(1 - e^{-t/\tau_{\text{on}}})} \times \tilde{h}_{\text{on}}(k_1 \tau_{\text{on}} + e^{-t/\tau_{\text{on}}}(z - k_1 \tau), 0), \quad (6)$$

where $\tau_{\text{on}} = (k_1 + k_3^{\text{on}})^{-1}$. $F_{\text{on}}(T)$ can then be computed for the different measurement scenarios using $\tilde{h}_{\text{on}}(x, 0) = G_s(x)/G_s(1)$ with $s =$ off for SCE and $s =$ on for SSE.

Peak in the distribution.—Our results, illustrated in Fig. 3, show a striking effect of orientational control for this model switch: for the SCE, we can obtain a peak in the flipping time distribution. Such a peak has been postulated for the *fim* switch [9], where it might imply that the switch tends not to leave the on state before it has had time to synthesize fimbriae [3,9,10]. Time spent in the on state may also influence recognition by the host immune system. This peak in $F_{\text{on}}^{\text{SCE}}$ is a consequence of the feedback between the switch state and the level of R_1. Once in the on state, the rate of on to off flipping increases with time as R_1 is produced. In contrast, for a Poissonian switch, the rate of flipping is constant in time. For a peak to occur, the slope of $F_{\text{on}}^{\text{SCE}}$ at the origin must be positive, which implies

$$k_2 - (k_4^{\text{on}})^2/k_3^{\text{on}} - (k_1 + 2k_4^{\text{on}})\langle n_0 \rangle - k_3^{\text{on}} \langle n_0^2 \rangle > 0, \quad (7)$$

where $\langle \ldots \rangle$ denotes an average using the weight W_{on}. The l.h.s. of (7) can be evaluated numerically using the exact result (4) to compute $\langle n_0 \rangle$ and $\langle n_0^2 \rangle$. We can then determine the regions of parameter space where a peak exists, as shown in the shaded region of Fig. 3 for the SCE. Our results show that the presence of a peak is favored by large values of k_2 (strong production of R_1 in the on state), and suppressed by very large values of k_3^{on} (strong R_1-mediated on to off switching) or by very small values of k_3^{on} (switching dominated by R_1-independent mechanism). Likewise, when k_4 is increased, the range of values over which the peak exists is decreased, since R_1-independent Poissonian

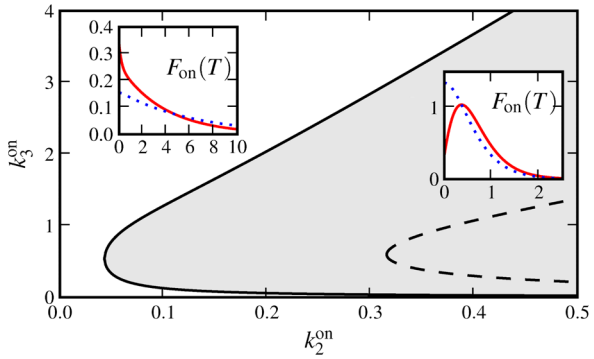

FIG. 3 (color online). Occurrence of a peak in the flipping time distribution $F_{\text{on}}^{\text{SCE}}(T)$, for $k_4 = 0.1$. In the shaded region the peak is present, and it vanishes outside this region. The dashed line shows the same result for $k_4 = 0.25$. As k_4 increases, the range of parameters for which there is a peak decreases. The unit of time is set by k_1 (i.e., $k_1 = 1$). The insets show examples of $F_{\text{on}}^{\text{SCE}}(T)$ (solid lines) and $F_{\text{on}}^{\text{SSE}}(T)$ (dotted lines). In the left inset, for $(k_2, k_3^{\text{on}}, k_4) = (0.2, 4, 0.1)$, $F_{\text{on}}^{\text{SCE}}(T)$ is peaked, while in the right inset, for (5,1,0.1), $F_{\text{on}}^{\text{SCE}}(T)$ shows no peak.

switching tends to dominate. For the SSE, on the other hand, we did not find any parameter values where inequality (7) is verified. The peak in the SCE appears because the number of recombinase, and hence the flipping probability, is typically low immediately after entering the on state and increases significantly thereafter. In contrast, in the SSE one typically starts a measurement when n, and hence the flipping probability, is already high. This tends to suppress the peak in the SSE flipping time distribution.

Correlated flips.—Another potentially important effect of the feedback between the switch state and the production of recombinase R_1 may be to cause correlations in the waiting times between successive flips (for example, a particularly short time before a flip might lead to subsequent flips occurring in quick succession). Such correlations might allow a population of switching microbes to ''remember'' the history of past environmental changes. We define the switch period τ_i as the time from when the switch enters the on state from the off state for the ith time, until it enters the on state for the $(i+1)$th time—i.e., the switch period is the duration of the on state plus the subsequent off state (cf. Fig. 1). In Fig. 4, we plot simulation results for the correlation function for switch periods τ_i and τ_j, as a function of $j-i$, as measured in simulations. For $k_3^{\text{off}} = 0$, weak correlation is observed between subsequent periods τ_i and τ_{i+1}. Correlations are weak because when $k_3^{\text{off}} = 0$, the off to on switching process does not depend on R_1 and is an uncorrelated Poisson process. When $k_3^{\text{off}} \neq 0$, correlations are much stronger and extended correlated sequences of flips emerge.

Discussion.—We have presented a generic model for a DNA-inversion switch with orientational control. By solving the model analytically in the case $k_3^{\text{off}} = 0$ (relevant to the *fim* switch), and using stochastic simulations, we have shown that this type of switch can display markedly non-Poissonian behavior, including a peaked flipping time distribution for intermediate values of k_3^{on} and, for $k_3^{\text{off}} > 0$, correlated sequences of flips. Non-Poissonian behavior has been postulated to be a consequence of orientational control [3,9]. The model presented here allows us to analyze the origins and effects of this behavior in detail, and provides analytical results which can be used as a basis for more complex models [13]. Suggested evolutionary roles for orientational control include rapid response to environmental change [10], as well as peaked flipping time distributions [3,9]. This study raises interesting questions about the consequences of non-Poissonian switching for population dynamics in changing environments. Several models have been proposed for the growth of populations of switching cells in stochastically and periodically changing environments (see, for example, [14]). These models assume Poissonian switch flipping. The analytical solutions presented here should make it possible to extend such models to the case of non-Poissonian flips. Non-Poissonian switch flipping opens up the possibility that lineages of cells may ''remember'' (in a statistical sense) the history of their recent phenotypic states. This is likely to have important consequences for models which include selection according to the fitness of different phenotypic states in changing environments. We speculate that bacteria which adopt a non-Poissonian flipping strategy may be able to maximize the evolutionary advantages of using some knowledge of the likely future behavior of the environment, combined with the benefits of a stochastic strategy as an insurance against sudden and unpredictable environmental changes. These avenues will be the subject of future work.

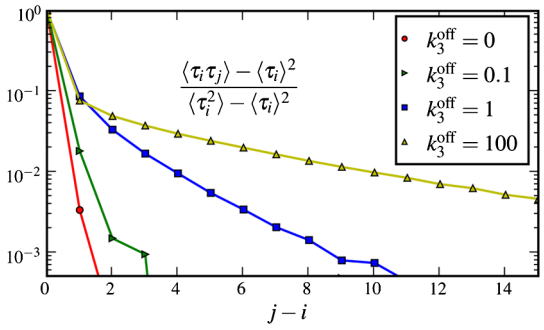

FIG. 4 (color online). Correlation function $[\langle\tau_i\tau_j\rangle - \langle\tau_i\rangle^2]/[\langle\tau_i^2\rangle - \langle\tau_i\rangle^2]$ between switch periods τ_i and τ_j, for several values of the rate k_3^{off}. These are simulation results for: $k_1 = 1$, $k_2^{\text{on}} = 5$, $k_3^{\text{on}} = 1$, $k_4^{\text{on}} = k_4^{\text{off}} = 0.1$.

The authors are grateful to A. Adiciptaningrum, G. Blakely, M. Cates, C. Dorman, A. Free, D. Gally, N. Holden, S. Tănase-Nicola, and S. Tans for valuable discussions. R. J. A. was funded by the RSE. This work was funded by the EPSRC under Grant EP/E030173.

[1] M. W. van der Woude and A. Bäumler, Clin. Microbio. Rev. **17**, 581 (2004).
[2] M. W. van der Woude, FEMS Microbiol. Lett. **254**, 190 (2006).
[3] I. C. Blomfield, Adv. Microbial Physiology **45**, 1 (2001).
[4] H. D. Kulasekara *et al.*, Mol. Microbiol. **31**, 1171 (1999).
[5] S. A. Joyce and C. J. Dorman, Mol. Microbiol. **45**, 1107 (2002).
[6] D. Dubnau and R. Losick, Mol. Microbiol. **61**, 564 (2006).
[7] J. L. Cherry and F. R. Adler, J. Theor. Biol. **203**, 117 (2000).
[8] P. B. Warren and P. R. ten Wolde, J. Phys. Chem. B **109**, 6812 (2005); Phys. Rev. Lett. **92**, 128101 (2004).
[9] D. W. Wolf and A. P. Arkin, OMICS **6**, 91 (2002).
[10] D. Chu and I. C. Blomfield, J. Theor. Biol. **244**, 541 (2007).
[11] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976); A. B. Bortz *et al.*, J. Comput. Phys. **17**, 10 (1975).
[12] Experiments are likely to be complicated by division events between flips (not considered here).
[13] N. J. Holden and D. L. Gally, J. Med. Microbio. **53**, 585 (2004).
[14] M. Thattai and A. van Oudenaarden, Genetics **167**, 523 (2004).